DOCUMENT RESUME

ED 136 551 FL 008 200

AUTHOR Dubuc, Robert
TITLE Banque de Terminologie de l'universite de Montreal. Description du systeme Termium (Terminology Bank of the University of Montreal. Description of the Termium System).
PUB DATE Jan 75
NOTE 22p.; In French

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Bilingualism; Computers; Data Bases; *Dictionaries; *English; *French; Information Sources; Information Storage; *Information Systems; Language Usage; Lexicography; Technical Writing; *Translation; *Vocabulary

ABSTRACT

The terminology bank at the University of Montreal is described. It consists of a dictionary of scientific and technical terms in the form of a bilingual card index permitting the entry of information into a computer and its retrieval via a terminal. The bank is intended for use by translators, government publication offices, information agencies, the media, publicity organizations, and technical and scientific research and education centers. Priority is given to new terms which cannot be found in traditional dictionaries or familiar terms which have acquired new meanings. The terminology bank is distinguished from a word bank in that the former is concerned with how the same scientific or technical words may be used differently in different languages. Thus, each entry provides a context of usage with a term, accompanied by the source and date of the context. (CLK)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED136551

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BANQUE DE TERMINOLOGIE
UNIVERSITE DE MONTREAL

TERMIUM

présente un dictionnaire de termes scientifiques et techniques, sous la forme d'un fichier bilingue à double contexte, réversible et complètement automatisé, permettant

- l'entrée des informations dans la mémoire d'un ordinateur, leur mise à jour ou leur consultation instantanées, par terminal,

- de façon à réunir et à tenir constamment à jour, en fonction des besoins de la clientèle à desservir, la plus grande masse possible d'informations terminologiques bilingues (ou multilingues).

L'information, emmagasinée sur pile CDC 844 de disques magnétiques, est accessible grâce à l'ordinateur CYBER 70, modèle 78, du Centre de Calcul de l'Université de Montréal.

Près de 55 000 fiches en mémoire (accessibles par plus de 134 000 entrées) peuvent déjà faire l'objet de consultations automatiques par télécommunication au moyen de terminaux imprimants ou de terminaux à écran, comportant (de préférence) un clavier à minuscules, majuscules et accents. L'information que contiennent ces fiches peut ainsi être obtenue instantanément à partir de la vedette anglaise, de la vedette française ou d'une partie de vedette. Elle peut également être obtenue par domaines, au moyen de l'interrogation directe par télécommunication ou séquentielle (en 24 heures).

Dès cette année, d'autres programmes offriront peu à peu de nouveaux accès à l'information en mémoire (par code de source, région d'utilisation et cote de pondération) ou assureront la mise à jour automatique instantanée du fichier par terminal.

La Banque entend desservir les bureaux de traduction et de rédaction des gouvernements et de l'entreprise, des agences d'information publiques ou privées, des organes de presse et de diffusion, des services de publicité et de relations publiques, les organismes de recherche ou d'enseignement technique ou scientifique, les études d'avocats, d'ingénieurs, etc. La Banque de terminologie répond présentement aux demandes terminologiques qui lui sont faites par téléphone. Son service de consultation automatique par terminal fonctionne expérimentalement.

Avec le temps, son fichier s'enrichira pour répondre à une proportion de plus en plus grande de demandes. Nos abonnés contribueront à hâter cet enrichissement car lorsque le terminal leur apprendra que le terme qu'ils demandent "n'est pas dans le fichier Termi", ils n'auront qu'à poser de nouveau la même question et, automatiquement, les terminologues de la Banque entreprendront sur-le-champ les recherches nécessaires pour trouver l'équivalent demandé.

Depuis 1972, le service de consultation téléphonique de la Banque a répondu en moins de 24 heures à plus de 90% des problèmes d'équivalences terminologiques anglo-françaises ou franco-anglaises qui lui ont été présentés.

FL008200

La Banque emmagasine toute l'information utile en donnant priorité à la mise en mémoire des termes de consommation courante sur lesquels les utilisateurs butent tous les jours: termes nouveaux qui n'ont pas encore pris place dans les dictionnaires traditionnels ou auxquels l'usage donne des acceptions nouvelles. Ce qui n'empêche pas la recherche thématique en divers domaines de se poursuivre constamment.

La Banque de terminologie n'entend pas normaliser. Elle tiendra compte de l'usage en y ajoutant une cote de pondération qui pourra rassurer, dans leur choix, le traducteur ou le rédacteur. Celui-ci trouvera fréquemment dans le code de source ou d'auteur un guide utile quant à l'emploi normal d'un terme dans un milieu dont l'identification transparaît sous le code.

La Banque a en mémoire automatique, le 21 février 1975:
63 296 fiches bilingues réparties en
22 classes (sur un total de 26)
268 divisions (sur 676)
et plus de
3 000 sections (sur 17,576) ou domaines d'application.
Ces 63 296 fiches sont automatiquement accessibles par
156 153 entrées ou termes (moyenne de 2,47 par fiche):
80 016 entrées en français
96 059 en anglais
et 78 en latin;
et par
101 773 indications de domaine (moyenne de 1,62 par fiche);
soit un total de 257 553 accès automatiques au fichier.

On peut donc, à l'heure actuelle interroger par télécommunication le fichier bilingue de la Banque de terminologie de l'Université de Montréal par plus d'un quart de million d'accès automatiques, sans compter les milliers de vocabulaires alphabétiques, bilingues ou multilingues, à partir du français, de l'anglais ou du latin, qui peuvent être demandés par télécommunication et expédiés dans les 24 heures. L'abonné demande le vocabulaire de son choix, dans le ou les domaines, ensemble ou séparément, qu'il désire.

Temps d'attente une fois la requête posée	3 secondes
Abonnement minimal (annuel) par terminal	$3,000.00
Temps de connexion (l'heure)	1.80
Tarif d'utilisation (par fiche) du terminal	0.05
Tarif minimum d'utilisation	1.00

BANQUE DE TERMINOLOGIE

de

L'UNIVERSITÉ DE MONTRÉAL

Description du systeme

T E R M I U M

par Robert Dubuc

Table des matières

		Page
0. Introduction		
0.1	Généralités	1
0.2	Le fichier Termium	1
0.3	L'exploitation informatique	2
1. Exploitation terminologique		
1.1	Les problèmes	4
1.2	Les solutions	4
1.3	Validité de l'information	5
1.4	Principes directeurs	5
1.5	Paramètres terminologiques	6
1.6	Classement des domaines	7
2. Fiche terminologique		
2.1	Contraintes	7
2.2	Contenu	8
2.3	Codes	9
2.4	Conception logique	11
2.5	Pseudo-fiche	12
2.6	Fiche de correction	12
3. Exploitation informatique		
3.1	Généralités	12
3.2	Contraintes	12
3.3	Objectifs	13
3.4	Système	13
3.5	Exploitation	14
3.6	Programmes d'exploitation directe	15
3.7	Programmes d'exploitation séquentielle	16
3.8	Fichiers de sécurité	16
3.9	Langage d'interrogation	16

édition révisée
Janvier 1975

0. INTRODUCTION

0.1. Généralités

0.1.1. La Banque de terminologie de l'Université de Montréal a été créée pour répondre aux besoins des usagers en remédiant à l'absence de documentation sûre, à sa dispersion et aux difficultés de consultation. A cette fin, elle s'est fixé un triple but: réunir la documentation terminologique présentement existante sous diverses formes et souvent difficilement accessible, susciter des efforts méthodiques de recherches terminologiques dans les secteurs non encore explorés et mettre toute l'information disponible à la disposition des utilisateurs, à prix raisonnable, selon des procédés rapides et efficaces.

0.1.1.1 Il importait dès lors pour l'orientation du projet de distinguer nettement entre une banque de mots et une banque de terminologie. La banque de mots vise une exploitation raisonnée des termes dans une langue donnée en facilitant le groupement des familles étymologiques, morphologiques ou sémantiques de façon à fournir aux usagers des moyens d'expression améliorés. La banque de terminologie, pour sa part, a essentiellement pour objet d'étudier le vocabulaire en situation; lorsqu'il s'agit de domaines scientifiques et techniques qu'il faut étudier en fonction de plusieurs langues, on s'aperçoit que les perspectives de travail dans l'un et l'autre cas sont loin d'être identiques. Dans le second cas, le problème de terminologie (définition du concept et équivalents dans les autres langues) devient en quelque sorte le noyau de l'information autour duquel viennent graviter les informations secondaires (paramètres terminologiques).

0.1.2 Pour réaliser ce projet, la Banque a décidé de créer un réservoir de base de fiches terminologiques bilingues, consignées dans une mémoire magnétique d'ordinateur, consultable directement au moyen de terminaux à clavier. Le système informatique régissant la réalisation de ce projet porte le nom de TERMIUM. Dans les pages qui vont suivre, nous allons tenter de le décrire et de préciser son orientation fondamentale, son contenu et son exploitation.

0.1.3 Le fichier TERMIUM implique une étape d'emmagasinement massif, qui s'effectue parallèlement aux étapes de la conception et de la mise au point du système informatique.

0.2. Le fichier TERMIUM

0.2.1. Le principe directeur qui a présidé à la conception du fichier est le suivant: fournir à l'usager des renseignements terminologiques suffisamment complets pour lui permettre de juger de leur portée. C'est pourquoi en principe chaque fiche donne l'information terminologique dans un contexte avec mention de source et de date.

0.2.1.1 En outre, en raison de notre situation géographique, économique et démographique, il a paru essentiel d'assurer le bilinguisme intégral du fichier, de façon à le rendre accessible à partir du français tout autant que de l'anglais.

0.2.1.2 Enfin, le trésor ou le corpus de TERMIUM constitue un fichier dont les renseignements ont une solide valeur documentaire mais ne constituent pas une norme en soi; les expressions normalisées l'ont été par des organismes extérieurs à la Banque et dûment mandatés à cette fin. En outre, TERMIUM ne prévoit pas que la lexicographie fasse partie de son exploitation principale. Il ne s'agit ni de mettre des définitions au point, ni de

publier des dictionnaires de spécialités, conçus dans l'optique des dictionnaires classiques. TERMIUM cherche à réunir des témoins de l'usage.

0.2.1.3 Cette orientation nous permet d'être constamment à l'écoute de l'usage, d'aiguiller la recherche en fonction des besoins, d'opérer rapidement des modifications et des rectifications et d'ajouter sans délai l'information récente.

Grâce à la technique d'accès direct et d'exploitation partagée, il sera possible de faire quotidiennement au fichier les corrections et les additions nécessaires, mettant ainsi rapidement l'information à la disposition des usagers.

Accès à l'information

0.2.2 On peut accéder à l'information de TERMIUM par les vedettes anglaises ou françaises de la fiche terminologique ou par l'un ou l'autre de leurs éléments déclarés importants. Par exemple, la fiche water-waste treatment plant, station d'épuration pourrait être consultée non seulement sous ces deux termes au complet, mais encore sous les mots water-waste, treatment, plant, station et épuration. Ce mode de consultation extrêmement souple permet l'exploitation maximale du capital terminologique, pallie le problème des usages synonymiques concurrents et peut mettre le consultant sur la piste d'une solution lorsque le fichier ne possède pas directement l'information demandée.

0.2.3. L'indication du domaine constitue une donnée terminologique essentielle, tant pour circonscrire une requête que pour étoffer l'information donnée. On sait que le mot anglais range peut avoir des significations fort différentes selon qu'il s'agit du domaine militaire, de l'électronique ou de l'acoustique. En plus d'ajouter une précision sémantique non négligeable, l'indication du domaine permet à la Banque de dresser par sujet divers répertoires terminologiques qui peuvent être d'une grande utilité pratique pour les rédacteurs, les lexicologues et les linguistes.

0.2.3.1 Le répertoire de domaines de la Banque possède une structure logique à trois niveaux: 26 classes, comportant 26 divisions, subdivisées en 26 sections, ce qui offre une capacité théorique de plus de 17,500 domaines.

0.2.4 Enfin, les mots sont aussi notés selon leur valeur. L'indice de pondération 0 signifie à proscrire, 1, douteux, 2, acceptable et 3, normalisé. Grâce à un traitement adéquat de programmation, il est possible, pour faciliter le travail de normalisation, de faire dresser automatiquement la liste des mots douteux appartenant au domaine de l'électrotechnique par exemple.

Tel est donc fondamentalement le fichier TERMIUM. Nous l'avons voulu modeste dans ses objectifs mais immédiatement axé sur les besoins du milieu à desservir.

0.3. Exploitation informatique

0.3.1. Pour répondre le plus efficacement possible aux besoins présents et futurs du milieu, les responsables de TERMIUM, après consultation des promoteurs d'expériences analogues en Europe, ont opté pour l'exploitation en accès direct, c'est-à-dire que le système permet, au moyen de terminaux à clavier, la consultation multiple, instantanée et conversationnelle du fichier terminologique.

0.3.1.1 Cette forme d'exploitation est fondée sur un langage explicite de déclaration des entrées, ou mots d'accès, à partir de mots-vedettes dont les éléments significatifs sont déclarés accessibles. De plus, les disparités graphiques sont résolues par référence à un mot bloqué d'où sont éliminées les sources de disparité (présence de traits d'union, majuscules, minuscules, etc.)

0.3.1.2 Pour que l'accès direct soit rentable, le système est conçu de façon à rendre le temps de consultation et de réponse indépendant en pratique du volume du fichier ou de la longueur de la réponse à fournir.

0.3.2 Bien qu'intégralement bilingue, le système n'exclut pas pour autant la possibilité d'introduire d'autres langues. Le mode de notation de langues permet l'adjonction d'un maximum de seize langues.

0.3.3. TERMIUM comprend fondamentalement trois fichiers essentiels: un fichier principal contenant les fiches-mères (vedettes et paramètres terminologiques) ainsi que les fiches-filles correspondantes (contextes ou définitions), un fichier où toutes les entrées sont classées par ordre alphabétique et casées par blocs et enfin un fichier index permettant de retrouver le bloc de classement du mot cherché.

0.3.3.1 Le système offre dans l'ensemble l'avantage de ne fermer aucune possibilité d'évolution et d'adaptation de façon à rester toujours au diapason tant du marché à desservir que du progrès technique.

0.3.4 TERMIUM prévoit en outre des fichiers auxiliaires d'exploitation dont les principaux sont la liste des sources, la liste des auteurs de fiches, l'index des domaines, les fichiers statistiques et bibliographiques et les fichiers de sécurité nécessaires à la préservation des informations stockées.

TERMIUM offre également diverses possibilités d'utilisation du fichier en séquentiel, en particulier pour les travaux de normalisation, de correction, de mise à jour et de lexicographie. Peu coûteux et commode, ce mode d'exploitation est appelé à rendre en pratique de grands services.

0.3.5 Enfin, il convient de rappeler que TERMIUM dispose d'un arsenal typographique complet avec majuscules, minuscules et accents. Bien qu'onéreuse, cette particularité a paru nécessaire en raison de la rigueur de traitement qu'exig l'information terminologique et aussi en raison des exigences d'une clientèle particulièrement sensible à la valeur des signes orthographiques.

0.3.6 L'accès direct est le mode privilégié de consultation, et accessoirement peut servir à l'emmagasinement et aux corrections. Toutefois, en raison du coût élevé et du contrôle complexe de l'entrée des informations en direct, une méthode de création massive a été mise au point à l'aide de l'appareil Flexowriter. Cet appareil permet dans une même opération de faire une copie codée et une copie en clair de la fiche terminologique, facilitant ainsi le contrôle et la correction immédiate des erreurs. Cette méthode permet l'entrée de 100 fiches par jour normal de travail par machine, norme qui répond au taux de croissance prévu pour le capital terminologique.

0.3.7. Les fichiers TERMIUM sont emmagasinés sur piles de disques 844 reliées à l'ordinateur CYBER 74 modèle 78 du Centre de calcul de l'Université de Montréal, jusqu'à concurrence de 100,000 fiches par pile.

0.3.7.1 Tout terminal à clavier bilingue de 94 caractères (ASCII) avec majuscules, minuscules et accents correspond aux exigences terminologiques de la Banque

et peut, quelle qu'en soit la marque, communiquer avec Termium. Parmi les principaux modèles imprimants ou à écran présentement commercialisés on trouve:

Olivetti (TE-318)
Lektromedia (LEK-110)

1. Exploitation terminologique

1.1. Les problèmes terminologiques qu'il importe de résoudre si l'on veut assurer l'efficacité de la communication dans les secteurs scientifiques et techniques sont les suivants:

1.1.1 Dispersion des efforts. Il se fait à l'heure actuelle bien des efforts de recherches dans diverses aires de vocabulaire mais, faute d'un organisme de coordination efficace, ces efforts dédoublent souvent des recherches déjà faites ou en cours, ou ne tiennent pas suffisamment compte des besoins prioritaires du milieu.

1.1.2 Valeur inégale des travaux. En plus du manque de coordination du travail, on s'aperçoit que les travaux terminologiques qui atteignent le stade de la publication sont d'une qualité fort aléatoire, faute d'une méthode rigoureuse de recherche terminologique et de critères établis justifiant la qualité de l'information.

1.1.3 Difficultés d'accès à l'information. Les travaux de terminologie n'ont souvent que de faibles tirages et deviennent très vite difficilement accessibles. En outre, la multiplicité des sources complique, de façon singulière, la consultation. Enfin, à cause de la rapidité d'évolution des sciences et des techniques, l'information terminologique dans les secteurs de pointe est pratiquement introuvable sous une forme immédiatement exploitable.

1.1.4 Fardeau de la recherche terminologique en traduction. L'interrelation des peuples accroît sans cesse le débit des traductions. Les bureaux de traduction doivent faire face chaque jour à un volume croissant de textes à traduire dont la complexité va également croissant. Dans l'état actuel des choses, au moins le tiers du temps du traducteur est consacré à la recherche terminologique. Ce qui est plus grave, c'est que ces recherches sont le plus souvent improductives parce que le traducteur ne consigne pas de façon systématique le résultat de ses recherches, de sorte que, le plus souvent, chacun fait des recherches déjà effectuées.

1.1.5 Lenteur des méthodes classiques de diffusion de l'information terminologique. En raison de la lenteur des mécanismes classiques d'édition, on peut dire qu'un dictionnaire, à peine publié, est déjà dépassé. De plus, la réponse doit souvent être fournie dans des délais très brefs, ce qui implique une rapidité d'accès à l'information tout à fait impossible à réaliser avec les fichiers artisanaux ou les ouvrages spécialisés classiques.

1.2. Les solutions

1.2.1 Sur le plan terminologique, la solution à ces divers problèmes repose sur une systématisation de la recherche terminologique, de façon à assurer la validité de l'information et le regroupement de tous les travaux faits dans les divers secteurs.

1.2.2 D'emblée le recours à l'informatique s'est imposé comme la seule solution capable de permettre la rapidité d'entrée des informations terminologiques,

leur regroupement global et selon divers critères d'utilisation, ainsi que leur tenue à jour constante et rapide.

1.2.3 En outre, en raison de l'évolution des problèmes terminologiques et de leur répercussion, en particulier sur la traduction, il a paru nécessaire de prévoir des mécanismes automatiques d'accès à l'information en offrant l'accès direct et conversationnel à l'usager au moyen d'un terminal à clavier dont les possibilités d'impression répondent aux critères de rigueur que les usagers réclament d'une banque de terminologie, c'est-à-dire présence des majuscules, des minuscules, des accents et autres signes diacritiques. Le recours à ces techniques peut permettre, en traduction en particulier, une accélération du débit de travail qui compenserait largement le prix de la consultation.

1.3 La validité de l'information.

Pour assurer à l'information régie par TERMIUM une validité qui, tout en étant suffisante aux yeux de la clientèle, réponde également aux critères observés par les organismes sérieux de terminologie de par le monde, la Banque a pris à son compte les principes de validité reconnus par le Congrès international de terminologie de Genève en décembre 1969, à savoir

a) que l'information terminologique doit toujours être citée avec contexte ou définition;

b) qu'elle doit être accompagnée de la mention de source et de date;

c) qu'elle doit être tirée de textes originaux dignes de foi.

1.4. Principes directeurs.

Ces principes une fois acceptés ne dispensent pas de la nécessité de répondre aux besoins de la clientèle virtuelle qui pourrait recourir à un fichier automatique.

1.4.1 La donnée de base qui conditionne toute l'orientation de la terminologie au Canada tient au caractère bilingue du pays. Les problèmes de terminologie se posent le plus souvent dans une optique de traduction, et de traduction de l'anglais (ou de l'américain) au français. Cette donnée suffit, croyons-nous, à justifier le caractère essentiellement bilingue du fichier TERMIUM.

1.4.2. Notion d'entrée.

Par entrée il faut entendre les termes qui donnent accès à l'information emmagasinée. Ainsi, pour la fiche qui figure à l'alinéa 2.2.3, les deux vedettes constituent chacune une entrée, soit waste-water treatment plant et station d'épuration; si le terminologue juge que les mots treatment et plant doivent également donner accès à cette fiche, nous aurons alors deux entrées supplémentaires, de même que nous aurons une troisième entrée supplémentaire si le terminologue estime que le mot épuration doit aussi donner accès à cette fiche. Le cas échéant, cette fiche comporterait cinq entrées.

1.4.2.1 Le langage de déclaration des entrées, c'est-à-dire les mots donnant accès à l'information terminologique, repose essentiellement sur le problème terminologique en relation avec les deux langues de travail fondamentales de la Banque, l'anglais et le français. Au moment de l'emmagasinement, les terminologues déterminent quels sont les mots d'accès. Ainsi, la

fiche velocity operated type microphone sera accessible à partir de l'anglais par la locution au complet et par microphone, puisqu'il est assez peu probable qu'on cherche à trouver la traduction de cette locution soit par velocity, operated ou type pris isolément ou même pris en bloc. La fiche sera encore accessible par l'expression française correspondante microphone à vitesse sans plus, parce que l'accès par microphone et vitesse isolément ne présente aucun intérêt terminologique. Il en est de même pour les unités complexes correspondant à une unité simple. Ainsi, pour la fiche bird's eye view - plongée, il est sans intérêt d'accéder aux éléments de l'unité complexe. Cette solution, même si elle présente le risque d'un jugement subjectif de la part des terminologues, élimine les accès inutiles, la surcharge de la mémoire et diminue les risques d'avalanche au moment de la consultation.

1.4.3 Le critère de sélection des contextes tient d'une part à leur valeur significative: il faut que le contexte permette au consultant de se faire au moins une certaine idée de la notion qui fait l'objet de la fiche. Un second critère tient à la parenté des contextes dans les deux langues. Cette parenté est souvent difficile à réaliser, mais, si on y parvient, la fiche offre alors des garanties maximales de sécurité. De plus, des contextes qui se recoupent offrent souvent matière à d'autres fiches, ce qui représente une économie appréciable de temps et de recherche. Ainsi, si, dans la définition de frequency response et dans celle de réponse de fréquence, on rencontre l'expression reference frequency - fréquence de référence, avec ces deux mêmes contextes, il est possible d'établir une nouvelle fiche sous cette nouvelle rubrique sans avoir à faire de recherches supplémentaires.

1.4.4. Les renseignements doivent être tirés de textes originaux dignes de foi, c'est-à-dire de textes qui ont la réputation d'être sérieusement rédigés. Les textes traduits, sauf en de rares exceptions, ne sont pas réputés pouvoir constituer une source valable d'information terminologique: les risques de contamination sont trop grands et très souvent le traducteur n'a pu ou n'a pas voulu prendre la peine de vérifier rigoureusement sa terminologie. Cette méthode offre, en outre, l'avantage de faciliter la découverte des tours idiomatiques propres à chaque langue de travail.

1.4.4.1 Chaque fiche doit porter la mention de source et la date afin de permettre à quiconque voulant vérifier ou approfondir l'observation de se reporter à l'ouvrage cité. En conséquence, la référence bibliographique doit être complète et la référence à la citation également (page, volume, etc.). La mention de la date apparaît nécessaire en raison de l'évolution rapide des techniques et des sciences. Il est sûr qu'une information en électronique, par exemple, datée de 1940, n'aurait pas la même valeur qu'une information datée de 1973.

1.5. Mais une banque de terminologie ne peut se dispenser de tenir compte de certains paramètres sémantiques ou grammaticaux qui ont effectivement une portée terminologique essentielle, d'où la nécessité de noter parfois le niveau de langage du terme employé (s'il s'agit d'un terme familier, argotique, s'il est pris dans un sens générique ou spécifique). Sur le plan grammatical, le genre et le nombre ont souvent une fonction sémantique particulière: il est alors normal d'en tenir compte.

1.5.1 En raison de l'orientation même de la Banque, qui s'écarte délibérément du travail lexicographique strict, pour une solution nettement terminologique plus près de l'usage et plus souple, il devient impérieux pour la sécurité du consultant d'évaluer la validité de l'information donnée. Une table

d'indices est prévue à cette fin.

1.6. Comme on prévoit un capital de roulement de la Banque supérieur à 250,000 fiches, il est nécessaire, pour faciliter la consultation et le regroupement de l'information selon des pôles rationnels, de préciser le ou les domaines d'emploi du terme étudié.

1.6.1 À cette fin, la Banque a mis au point un répertoire de classification des domaines qui possède une structure logique à trois niveaux: 26 classes forment l'armature du classement et groupent les domaines fondamentaux d'activités qui se prêtent à un développement terminologique. Ces 26 classes comportent 26 divisions qui représentent à l'intérieur d'une même classe les principales aires d'activité, et les divisions se divisent à leur tour en 26 sections représentant autant de situations terminologiques caractérisées par un centre d'activité. Ainsi, dans le code AFF, la première lettre représente la classe Administration et gestion, la seconde, la division Gestion du personnel et la troisième, la section Perfectionnement et formation du personnel.

Ce mode de classement offre une possibilité théorique de 17,500 sections de vocabulaire. Il offre, sur les codes standardisés déjà existants (classification décimale universelle, Dewey, Library of Congress), l'avantage d'être conçu exclusivement pour l'exploitation terminologique. À cette fin, il présente toute la souplesse nécessaire.

2. Fiche terminologique

2.1. La fiche terminologique doit, dans sa conception, répondre à des exigences de deux ordres. D'abord servir de véhicule à une information terminologique sûre et ensuite permettre de communiquer, sous une forme concise, cette information aux clients selon des modalités conformes à leurs besoins.

2.1.1. Pour répondre à la première exigence, la fiche terminologique prévoit pour chaque vedette la mention de la source, l'année de publication et la mention de page. À ces informations s'ajoutent les paramètres terminologiques portant les mentions de niveau ou de particularismes langagiers, la notation de langue, l'indice de pondération et les notes grammaticales concernant la nature de l'expression, le genre et le nombre, lorsque ces indications ont une valeur terminologique.

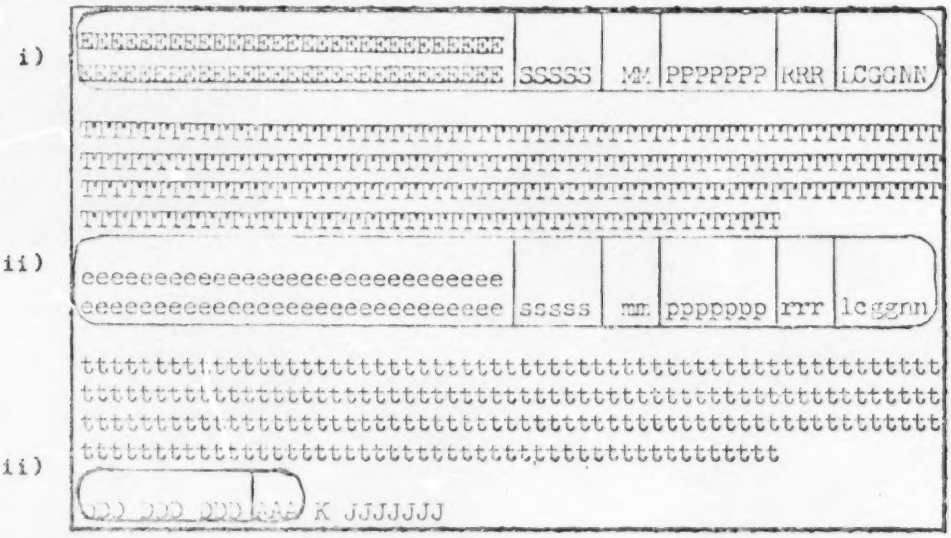
2.1.1.1 Les contextes ou définitions sont d'autre part consignés dans des modules de 225 caractères. L'expérience prouve que cette taille permet,dans la très grande majorité des cas,de retenir en un seul module l'essence d'un contexte ou d'une définition.

2.1.2 Enfin, trois indications s'appliquent à l'ensemble de la fiche, soit les domaines d'application, donnés en code, jusqu'à concurrence de trois, l'indicatif d'auteur de la fiche, attestant la propriété intellectuelle de celui qui établit la fiche, et enfin l'indice de pondération globale qui n'évalue que l'équivalence des vedettes.

2.1.3 La fiche Termium tient compte à peu près de toutes les informations nécessaires pour attester la validité de l'information recueillie. Par son format, la fiche 3 x 5 po. (75 x 125 mm) impose une certaine rigueur qui force à l'essentiel, évitant ainsi l'écueil des informations parasites. Ce format s'intègre en outre à la majorité des fichiers existants au Canada et est conforme au format standard international.

2.1.4 Il importe aussi de disposer les renseignements de façon claire et ordonnée pour que l'usager puisse facilement se retrouver. La division adoptée pour la fiche permet de rattacher, sans aucune ambiguïté possible, tous les renseignements aux vedettes pertinentes. De même, les renseignements portant sur l'ensemble de la fiche sont nettement dégagés du reste. Par sa forme synthétique, la fiche se présente donc comme un véhicule idéal de diffusion de l'information terminologique.

2.2. Contenu de la fiche


i)
EEEEEEEEEEEEEEEEEEEEEEEEEEEE
eeeeeeeeeeeeeeeeeeeeeeeeeeee
SSSSS
MM
PPPPPPP
RRR
LCGGMN
TTTT

L,l: Langue de la vedette (1 car.)

C,c: Pondération de la vedette (1 car.)

G,g: Catégorie grammaticale (1 ou 2 car.). Non indispensable. Verbe, adjectif, adverbe, nom, etc.

N,n: Genre et nombre (1 car. chacun). Non indispensable.

T,t: Définition ou contexte de la vedette (max. de 225 car.).

2.2.3 C. Contenu de la zone 3

D: Code des domaines (maximum de 3 domaines).

A: Auteur de la fiche (indicatif fourni par la Banque) (3 car.).

K: Pondération de l'équivalence des vedettes (1 car.)

#: À moins d'indications explicites contraires, le système informatique adopte une notation uniforme préétablie, dite option de défaut.

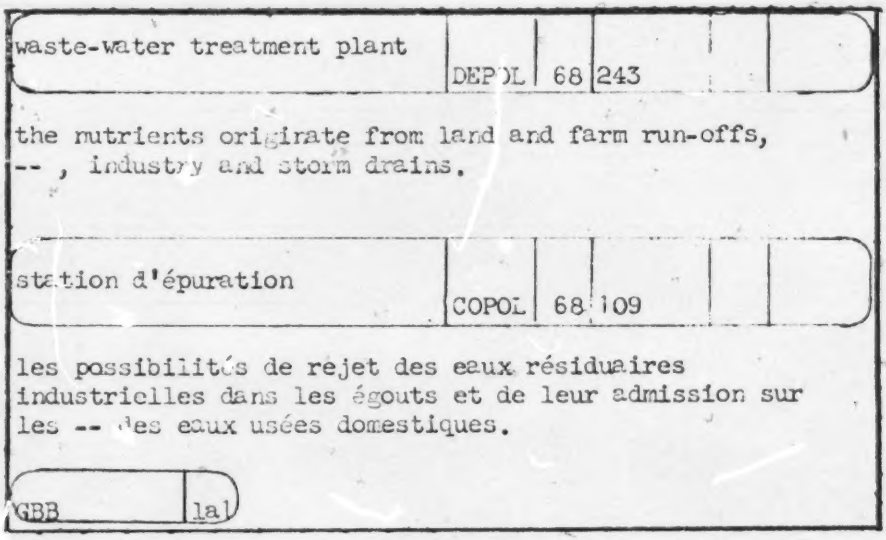

waste-water treatment plant
DEPOL 68 243
the nutrients originate from land and farm run-offs,
-- , industry and storm drains.
station d'épuration
COPOL 68 109
les possibilités de rejet des eaux résiduaires
industrielles dans les égouts et de leur admission sur
les -- des eaux usées domestiques.
GBB
lal

Fig. 2. - La fiche remplie

2.3. Les codes

2.3.1 L'adoption d'un format réduit pour la fiche rend nécessaire l'utilisation de codes pour les sources, les particularismes, la pondération, les notes grammaticales, les domaines et les auteurs de fiches.

2.3.2. Ces codes en général sont simples et se rapprochent le plus possible de l'information explicite.

2.3.2.1 Le code de source est formé d'un sigle prononçable de cinq caractères, combinaison du nom de l'auteur et du titre. Le code des noms de revues est habituellement formé par l'abréviation du titre de façon à former un

mot prononçable, pour qu'il soit plus facile à retenir. Dans chaque cas, le code renvoie à une fiche bibliographique complète.

2.3.2.2 Le code des particularismes veut identifier l'aire d'utilisation géographique du terme. Les principaux symboles sont:

AE: anglais australien
CE: anglais canadien
FA: français africain
FB: français belge
FC: français canadien
FF: français de France
FR: français régional
FS: français suisse
FU: français universel
GB: anglais britannique
RE: anglais régional
US: anglais américain

À ce code s'ajoutent les précisions d'aire sémantique, sociale, chronologique, didactique ou de fréquence:

A: archaïque
B: propre
C: courant (opposé à savant)
D: dépréciatif ou péjoratif
E: rare, exceptionnel
F: niveau familier
G: générique
I: adjectif d'inhérence
J: jargon technique
K: figuré
L: niveau littéraire
M: marque commerciale
N: néologisme
P: niveau populaire ou argotique
R: adjectif de relation
S: spécifique
T: savant ou technique (opposé à co
V: vieilli
Y: concret
Z. abstrait

2.3.2.3 Code des langues

1. anglais 2. français 3. latin 4. allemand

Pour le français et l'anglais, cette mention ne se fait que dans les pseudo-fiches, le système informatique identifiant automatiquement ces langues sur la fiche de base.

Cette notation permet l'adjonction éventuelle d'autres langues.

2.3.2.4 Code de pondération pour chaque vedette et pour la fiche dans son entier.

0: à proscrire
1: douteux
2: acceptable (noté automatiquement par le système)
3: normalisé

2.3.2.5 Code de la catégorie grammaticale

N (plus espace):	nom commun	V (plus espace):	verbe
N:	nom propre	VI:	verbe intransi
AJ:	adjectif	VT:	verbe transiti
AV:	adverbe		

2.3.2.6 Genre et nombre

M: masculin
F: féminin
X: neutre
S: singulier
P: pluriel

Les codes prévus en 2.3.2.2, 2.3.2.5, 2.3.2.6 ne sont donnés que si ces indications ont une portée terminologique.

2.3.2.7 Code des domaines, déjà expliqué en 1.6.1.

Lorsqu'on indique plus d'un domaine, chaque indication est séparée par un espace. On peut se procurer le Répertoire en s'adressant à la Banque.

2.3.2.8. Code d'auteur de fiches

2.3.2.8.1 Les auteurs de fiches peuvent être soit des organismes, soit des particuliers

2.3.2.8.2 S'il s'agit d'organismes, la Banque attribue à chacun le code de deux lettres en majuscules (aussi près que possible du sigle déjà en usage par l'organisme visé), plus indicatif chronologique (numérique), s'il y a lieu.

Exemples:
CN: Chemins de fer nationaux
RC: Radio-Canada
CP: Canadian Pacific
CI: Canadian Industries Limited
DT: Domtar

2.3.2.8.3 S'il s'agit de particuliers, le code est formé de leurs initiales en minuscules (2 lettres plus indicatif chronologique (numérique). Lorsque les dix unités numériques (1 à 0) auront été utilisées, on remplacera le chiffre par les lettres majuscules (A à Z) ou on utilisera d'autres combinaisons de ces lettres.

Exemples:
rd (1 à 0)
rd (A à Z)
Rd (1 à 0)
Rd (A à Z)
rD (1 à 0)
rD (A à Z)
RD (1 à 0)
RD (A à Z)

2.3.2.8.4 Chaque auteur de fiche agréé par la Banque s'inscrit en donnant les renseignements suivants: Nom, prénom usuel, profession, adresse personnelle, adresse professionnelle, numéro de téléphone, nombre d'années d'exercice de la profession indiquée, spécialités.

2.4. Conception logique de la fiche

Les informations contenues dans la fiche se divisent en deux parties.

2.4.1 La première (fiche-mère) comprend les données terminologiques essentielles, c'est-à-dire les vedettes correspondantes et leurs paramètres, auxquels s'ajoutent la mention des domaines d'application, l'indicatif d'auteur de la fiche et l'indice de pondération globale.

2.4.2 La seconde (fiche-fille) comprend pour chaque langue les exemples d'utilisation (contextes) ou des définitions. Ces renseignements n'ayant qu'une valeur documentaire par rapport à ceux de la première partie, l'utilisateur peut éviter de surcharger la réponse à sa consultation en ne demandant que des éléments de la fiche-mère.

2.4.3 Les informations essentielles sont donc, pour la mise en mémoire, groupées dans une fiche-mère qu'on pourrait représenter graphiquement ainsi:

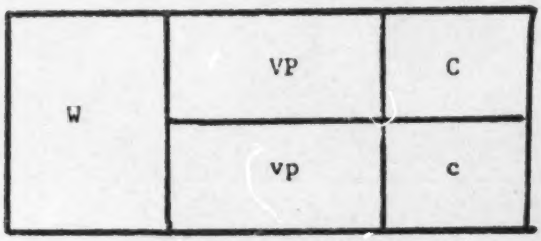


fig. 3

W: Données communes aux deux vedettes
VP: Vedette et paramètres de la première langue
vp: Vedette et paramètres de la deuxième langue
C: Pointeur pour repérer le contexte de la première langue
c: Pointeur pour repérer le contexte de la deuxième langue

2.4.4 Les informations contextuelles sont consignées sur des fiches-filles repérables par les pointeurs de la fiche-mère. Ces informations peuvent prendre la forme d'un simple contexte, d'une définition, d'une observation grammaticale, d'une explication sémantique, etc.

2.5. Pseudo-fiche

2.5.1 La pseudo-fiche offre un moyen d'élargir, en cas de besoin, les cadres de la fiche fondamentale lors de l'entrée en mémoire. Elle peut servir à inscrire des locutions à l'intérieur des vedettes de la fiche-mère, à ajouter des éléments d'informations, à augmenter le nombre des domaines d'application. Dans le langage scientifique, elle permet aussi de tenir compte des appellations latines ou allemandes.

2.5.2 Chaque pseudo-fiche porte un indicatif spécial PF et suit obligatoirement la fiche de base qu'elle prolonge. Elle s'insère dans le système selon la nature de l'information qu'elle porte.

2.6 Fiche de correction

Tout comme la pseudo-fiche, la fiche de correction est un moyen de modifier l'information déjà en mémoire. Elle s'insère elle aussi dans la structure de base du système fiche-mère - fiche-fille, selon la nature de l'information qu'elle contient.

3.0 Exploitation informatique

3.1 Généralités

La fiche terminologique mise au point, il s'agit maintenant de voir comment le système d'exploitation informatique peut atteindre, en dépit de ses contraintes inhérentes, les objectifs fixés.

3.2. Contraintes

3.2.1 La première contrainte tient évidemment à l'ampleur du projet. Etablir un fichier bilingue pour les domaines scientifiques et techniques à une époque de découvertes et d'évolution accélérées pose des problèmes quantitatifs

d'emmagasinement de l'information. Il importait donc au premier chef de mettre au point un système économe qui évite l'écueil de l'engorgement et des embouteillages causés par une masse d'information qu'on ne peut plus raisonnablement manier.

3.2.2 À cette contrainte quantitative s'ajoute une contrainte qualitative; en raison du rôle important de diffuseur d'informations qu'une banque est appelée à remplir, il importe que l'information soit toujours le plus au point possible et ait toutes les qualités formelles que les usagers sont en droit d'attendre. Il fallait donc mettre au point un système facile à tenir à jour, permettant d'opérer avec un minimum de transformations toutes corrections utiles, et capable d'offrir les informations dans une forme qui réponde aux exigences de la clientèle. Il est certain qu'en français, par exemple, les clients accepteraient difficilement, à cause des répercussions sur la qualité de l'information, une fiche sans accents et sans majuscules. Or, les systèmes de télécommunication avec l'ordinateur ne disposent généralement pas en Amérique du Nord de l'arsenal typographique nécessaire pour transmettre l'information en français avec tous les signes diacritiques.

3.3. Objectifs

3.3.1 Termium a donc voulu, tout en respectant ces contraintes, atteindre les objectifs qui lui semblaient essentiels pour assurer le fonctionnement efficace d'une banque de terminologie.

3.3.2 Il fallait d'abord concevoir une technique d'emmagasinement rapide, simple et peu coûteuse, qui assure une qualité maximale du produit, permette des corrections avec un minimum de transformations et puisse servir également, dans les mêmes conditions, aux mises à jour massives.

3.3.3 Le système doit également être rapide à la sortie, c'est-à-dire qu'il doit permettre l'accès à l'information emmagasinée dans des délais minimaux, en raison des besoins de la clientèle, du coût élevé d'utilisation du temps machine et de la masse énorme d'informations stockées.

3.3.4 Le système doit en outre prévoir des utilisations auxiliaires ou dérivées du capital stocké. Permettre, par exemple, le regroupement de l'information par domaine, par cote de pondération, par auteur de fiche, etc.

3.3.5 De même pour orienter toujours le système conformément aux besoins du marché, il est nécessaire d'établir certaines statistiques sur la nature et l'objet des requêtes faites par les clients.

3.4. À la lumière de ces objectifs et contraintes, les informaticiens du Centre de calcul ont mis au point un système vraiment sur mesure pour assurer à la Banque de terminologie de l'Université de Montréal une efficacité maximale.

3.4.1. Création massive

La création massive, c'est-à-dire l'emmagasinement en masse de l'information terminologique, se fait au moyen de l'appareil Flexowriter qui permet d'établir la fiche en clair en même temps que la fiche codée sur ruban perforé. Ce système offre l'avantage d'être rapide; une perfodactylo d'expérience peut facilement taper cent fiches complètes par jour de travail. La copie de la fiche étant également établie en clair, le repérage des erreurs peut se faire immédiatement et les mécanismes de corrections s'en trouvent considérablement simplifiés. Le recours au

Flexowriter répond donc très bien aux contraintes et aux objectifs concernant la mise en mémoire massive de l'information.

3.4.1.1 L'utilisation du Flexowriter implique la mise au point de trois programmes: un programme de décodage de la bande-papier perforée; un programme régissant l'ingestion et la digestion des informations de la bande qui donne, sur imprimante rapide, la liste des entrées déclarées, la liste des fiches stockées, leur numéro d'identité et l'indication d'acceptation ou de rejet; enfin, un programme de correction permettant de faire les modifications indiquées par la liste sur imprimante rapide.

3.4.1.2 Mises à jour et créations exceptionnelles

La mise au point du mode de création massive n'exclut pas la possibilité d'introduire en direct,au moyen du terminal, des corrections, des additions ou des soustractions de caractère urgent. Toutefois, en raison des avantages de sécurité et d'économie de l'emmagasinement par Flexowriter, cette possibilité demeure exceptionnelle.

3.5. L'exploitation

3.5.1 Pour répondre aux besoins des clients et pour tenir compte de l'évolution des techniques informatiques, la Banque a fait de TERMIUM un système de consultation en accès direct et conversationnel.

3.5.2 En raison du volume de l'information et du service à assurer aux clients, il était essentiel que le temps de réponse de l'ordinateur soit pratiquement indépendant du volume de la Banque ou de la longueur de la requête. Le système TERMIUM a permis de résoudre cet épineux problème par la conception astucieuse de ses fichiers d'exploitation.

3.5.3. Trois fichiers de base

La conception logique de la fiche terminologique a permis de mettre au point un système comportant trois fichiers de base.

3.5.3.1 Premier fichier

Il comprend, dans un premier volet, les fiches-mères contenant chacune une paire de vedettes et leurs paramètres ainsi que l'attribution des domaines, le code d'auteur de fiche et l'indice de pondération globale. Les fiches-filles contenant les contextes sont groupées dans un second volet. Pour chaque vedette des fiches-mères, un pointeur permet de récupérer les contextes correspondants. Chaque fiche-mère comporte en outre une adresse numérique de repérage.

3.5.3.2 Second fichier

Toutes les entrées déclarées (1.4.2) y sont groupées par ordre alphabétique à l'intérieur de blocs dûment identifiés. À chacune de ces entrées correspond l'adresse des fiches-mères correspondantes.

3.5.3.3 Troisième fichier

Un troisième fichier alphabétique où figurent les 7.5 premiers caractères de la première entrée de chaque bloc composant (adresse du bloc) permet l'accès au second fichier. La consultation procède donc d'abord par le balayage du fichier index (troisième) qui donne l'indicatif du bloc (second)

où se trouve consignée l'adresse de la fiche-mère (premier) donnant la réponse à l'information cherchée.

3.6. Programmes d'exploitation en direct

Ce mode d'exploitation implique la mise au point d'une programmerie particulière.

3.6.1. Le programme de consultation en direct prévoit:

a) la vérification comptable pour établir l'identité du consultant et l'étendue de ses prérogatives;

b) l'analyse de la demande permettant à l'ordinateur de l'accepter, de la rejeter ou de transmettre, le cas échéant, un avertissement d'erreur;

c) l'exécution des demandes par la consultation automatique des fichiers avec acheminement de la réponse.

3.6.1.1 Conséquence de l'exploitation en accès direct, la Banque doit être en mesure d'évaluer les besoins par la compilation des statistiques de consultation, à savoir combien de fois chaque terme a été demandé, quels termes ont été demandés sans obtenir de réponse et avec quelle fréquence, etc. Le système doit pouvoir aussi établir automatiquement le nombre et la durée des requêtes par client.

3.6.1.2 Ces opérations sont assurées par des programmes gérant un autre fichier à deux volets qui donne d'une part le code comptable de chaque abonné avec le relevé de ses activités pour une période de référence donnée, et, d'autre part, le relevé ou le bilan d'activité de tout le système, ce qui peut permettre à la Banque de mieux connaître ses déficiences et les besoins du marché.

3.6.2. Mises à jour

Afin de pouvoir toujours se tenir au diapason de l'évolution des terminologies la Banque dispose de mécanismes de mises à jour en direct de son fichier. En plus d'assurer le stockage rapide de l'information, ce système permet d'apporter des corrections à la Banque d'une façon simple et efficace. Pour y arriver, un programme spécial permet:

3.6.2.1 de corriger (éliminer, ajouter, soustraire, modifier) sur la fiche tous les éléments (sauf les vedettes et les entrées supplémentaires);

3.6.2.2 de mettre à jour, par une méthode simplifiée, les indices de pondération tant pour chacune des vedettes que pour l'ensemble de la fiche;

3.6.2.3 d'opérer les modifications comptables exigées par

(i) l'introduction d'un nouvel usager;

(ii) la modification des prérogatives d'un usager à l'égard de la Banque (consultation et création, consultation seulement);

(iii) l'élimination d'un usager

3.6.3 L'exploitation en direct exige en outre l'établissement d'un fichier accessoire qui réunisse la liste des sources d'information terminologique

par code et par titre (avec référence bibliographique), la liste des auteurs de fiches par code et par nom; enfin, le catalogue et l'index des domaines visant à faciliter le classement et le repérage des termes.

3.7. Programmes d'exploitation en séquentiel

Conçu prioritairement pour la consultation en accès direct, le système TERMIUM n'exclut pas l'exploitation en séquentiel, car souvent, lorsque l'urgence n'exige pas le recours à un mode ultrarapide de consultation, le traitement en séquentiel (consultation massive) peut se révéler satisfaisant.

3.7.1 TERMIUM recourt à l'accès séquentiel pour faire faire la facturation aux abonnés, faire dresser les listes de fiches entrées en direct avec l'indice de pondération 1, faire tirer des bandes-flexo afin d'obtenir des copies en clair des fiches entrées en direct.

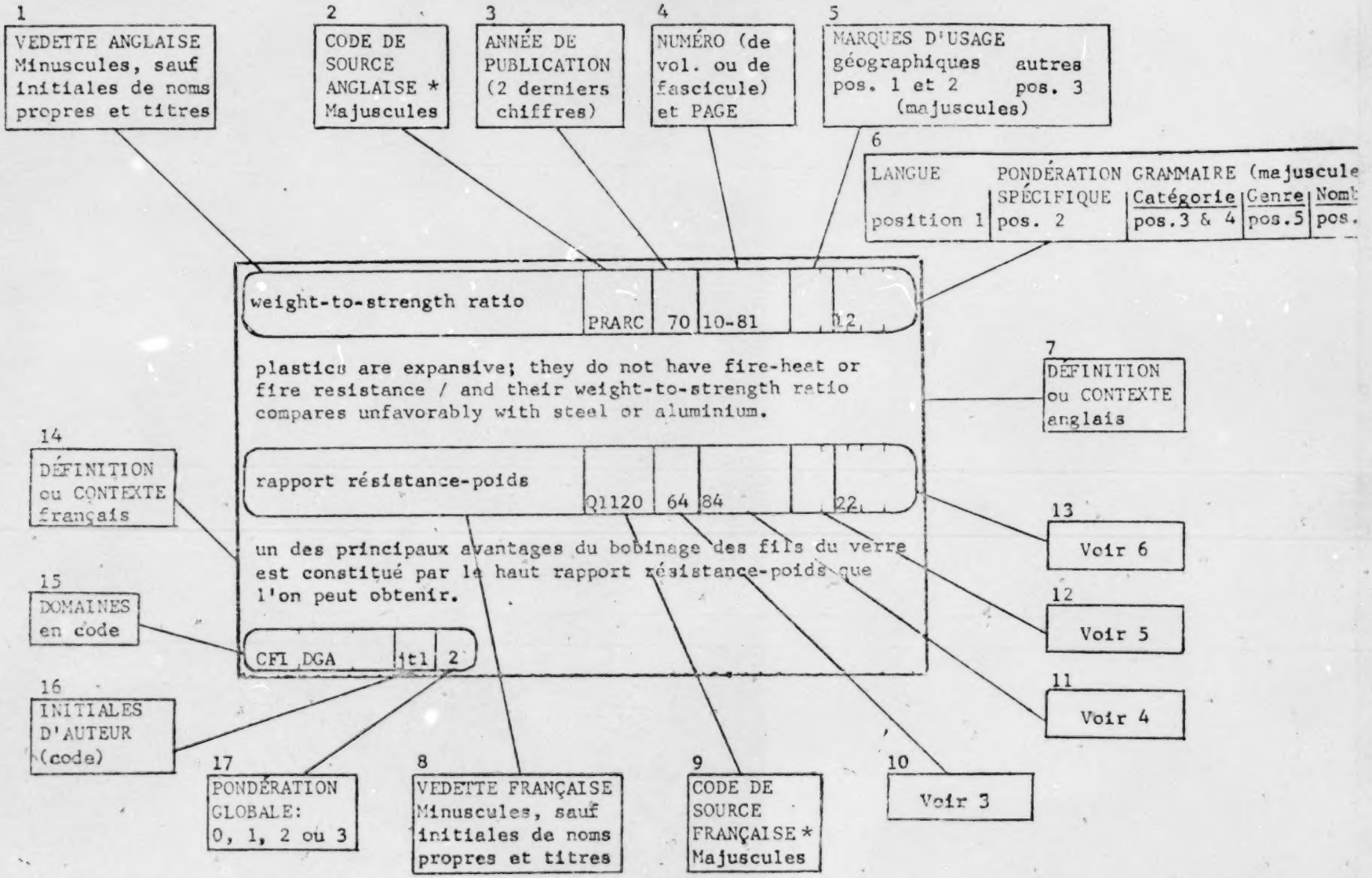
3.7.2 De même, les listes alphabétiques suivantes seront accessibles en séquentiel: les termes par domaine (classe, division ou section), les fiches d'un même auteur, les fiches selon leur indice de pondération, tout le capital terminologique de la Banque. Ces éléments peuvent aussi se combiner. On pourrait par exemple faire dresser la liste alphabétique des termes d'électroacoustique portant l'indice de pondération 1.

3.8 Pour parer à toute éventualité, le système doit prévoir des fichiers de sécurité qui assurent la conservation d'une copie intégrale sur disques des fichiers fondamentaux de la Banque (3.5.3.1, 3.5.3.2, 3.5.3.3) tenus à jour et la reproduction de ces disques sur bandes magnétiques. Ces mesures réduisent au minimum les risques de pertes accidentelles d'information.

3.9. Le système dispose d'un langage d'interrogation qui lui est propre et qui facilitera au maximum l'accès du client à l'information en le guidant graduellement vers la réponse à ses problèmes terminologiques. Un langage d'interrogation insuffisamment discriminant pourrait être coûteux, inefficace en plus d'être très frustrant pour l'usager. Ce dernier, à la demande du système, peut être appelé à préciser sa requête par l'identification du domaine d'application de façon à éviter d'être englouti sous une masse d'informations inutiles.

3.9.1 En ce qui concerne les corrections, en plus de la méthode d'accès direct déjà exposée, TERMIUM prévoit un langage spécial de correction, apparenté au mode d'emmagasinement massif des informations avec les mêmes avantages de sûreté et d'économie.

Schéma explicatif de la fiche réglementaire



* Dans les champs 2 et 9, le signe "@" représenterait le code de la Banque de terminologie.